Teoría de las comunicaciones

Primer cuatrimestre de 2017

Departamento de Computación
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Trabajo práctico 1: Wiretapping

Integrante	LU	Correo electrónico
Alejandro Albertini	924/12	ale.dc@hotmail.com

Reservado para la cátedra

Instancia	Docente	Nota
Primera entrega		
Segunda entrega		

Comentarios del corrector:

Índice

Índice de figuras

1. Introducción

En este informe nos proponemos analizar diferentes redes locales utilizando herramientas derivadas de la teoría de la información. El objetivo es usar y analizar datos de los paquetes que pasan por la red para descubrir patrones y topologias sin conocer previamente nada sobre los nodos de la red. Para lograr esto utilizamos direcciones de la capa de enlace en un principio y luego datos de paquetes `ARP` (address resolution protocol, un protocolo utilizado para obtener una dirección de enlace dada una dirección de red para que dos nodos en una misma red local de acceso compartido puedan comunicarse) para obtener mediciones de información y entropia de la misma red modelada de modo diferente. Además, se realizarán distintos tipos de gráficos, con las mediciones obtenidas y las fuentes propuestas, para tratar de mostrar y detallar los resultados obtenidos. Se mostrarán distintos tipos de gráficos, como la composición de una red y sus nodos o la cantidad de paquetes broadcast vs unicast, y se analizará el por qué de esos resultados obtenidos.

2. Métodos y ambientes

Las redes serán analizadas con dos herramientas que escuchan un medio compartido en modo promiscuo.

La primera herramienta modela la red como una fuente S cuyos posibles símbolos son $S_{unicast}$ si el paquete leído es enviado a un nodo específico y $S_{broadcast}$ si el paquete es enviado a la dirección broadcast (`ff:ff:ff:ff:ff:ff`). Además, ningún paquete es filtrado y cualquier paquete que llega nuestro enlace será tenido en cuenta. El objetivo en este caso es diferenciar cuando se envía un paquete a un sólo dispositivo (paquete unicast) o cuando se manda un mismo paquete a todos los dispositivos que están conectados a la red en ese momento (paquete broadcast).

La segunda herramienta modelará la red como una fuente $S1$ cuyos posibles símbolos son todas las direcciones IP disponibles en la red (como esto es desconocido, vamos a asumir que todos los símbolos observados son todos los símbolos posibles) en paquetes ARP que se van capturando en la red. Además, se agruparán los nodos iguales para mostrar un gráfico más entendible. El objetivo de ésta fuente es poder distinguir los distintos hosts que hay en la red, poder analizar como están conectados entre sí, descubrir el alcance total de la red y ver que dispositivos están conectados en la misma.

La consigna era definir una función que designe algún símbolo como distinguido fundado en algún resultado matemático. Por eso, distinguimos a aquellos símbolos que tengan información menor a la entropía, dado que estos serán más comunes de ver (por la definición de información y de entropía), ya que los otros nodos de la red piden su dirección de enlace muy frecuentemente. Esto hará distinguir a nodos internos muy usados o bien al gateway por defecto de la red, por la cual, todos los nodos salen a otras redes como internet.

El tiempo de muestreo es similar para todos los experimentos, fue de 30 minutos en cada caso.

Las 4 redes que elegimos para experimentar son las siguientes:

- Red hogareña: red local wi-fi en una casa particular, lo único que sabemos es que contiene una cantidad muy limitada de computadoras personales y dispositivos móviles.
- Red laboral: red local ethernet en una oficina, sabemos que conviven computadoras de empleados y servidores.
- Red pública: red wi-fi en un shopping, diferentes tipos de usuarios, dispositivos y duración de leasing the direcciones de red
- Red hogareña nuevo experimento: Para el recupetario del TP1, se decidió agregar ésta nueva red de hogar, mucho más pequeña que la primera red hogareña y en un domicilio diferente.

3. Red hogareña

Esta red en principio es muy controlada y sólo sabemos que cuenta con un número pequeño de nodos. La red usa en su mayoria wi-fi por lo cual es de esperar que el nodo que usamos para monitorear la red reciba amplia cantidad de paquetes who-has y is-at dado que no hay switches segmentando la red que filtren respuestas is-at de terceros.

El primer experimento, el cual ve la red como una fuente con 2 símbolos unicamente, arroja luego de monitorear la red, que la mayor parte del tráfico es unicast y solo el 17 % es broadcast.

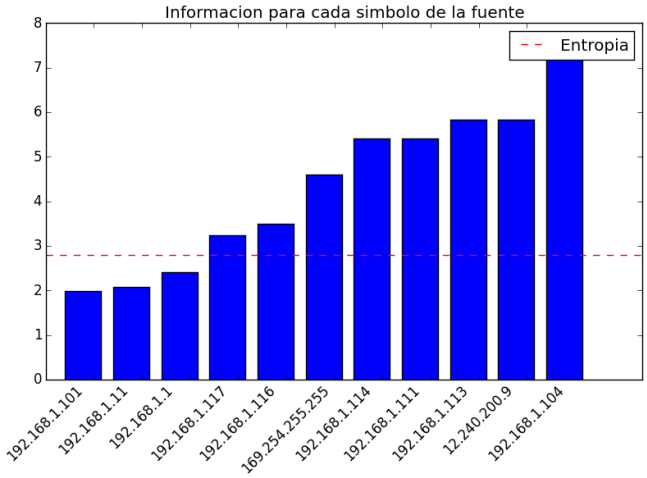
Figura 1: Información de S, Red hogareña

	frecuencia	información
$S_{broadcast}$	0.17	2.56
$S_{unicast}$	0.83	0.27

Estos valores nos dan una entropía de 0.66 bits (siendo el máximo 1 dado que hay 2 simbolos en principio equiprobables) lo cual concuerda con las expectativas dado que en una red local hogareña se espera ver mas que nada trafico unicast entre los nodos y el gateway hacia internet.

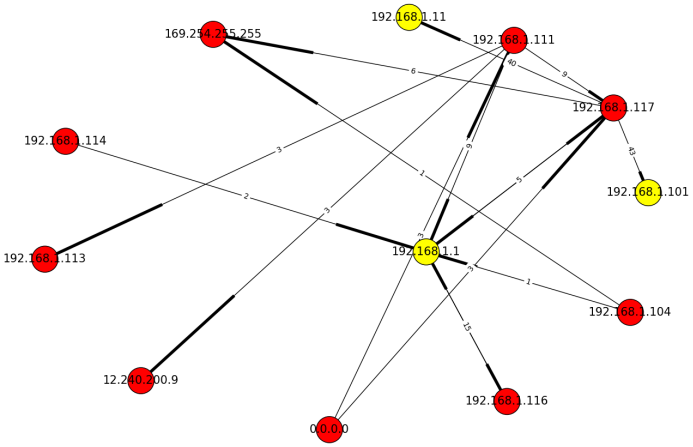
En el segundo experimento, el cual modela la red basado en la dirección a resolver en mensajes ARP, usamos la misma captura utilizada en el experimento anterior. El script que usamos para analizar los resultados destacó 3 nodos de la red cuando esperabamos sólo 1. El siguiente gráfico muestra la información de cada símbolo y la entropía de la fuente, como se puede ver hay 3 nodos destacados bajo nuestra definición.

Figura 2: Información red hogareña



Una investigación más meticulosa concluyó que `192.168.1.1` es el gateway por defecto de la red mientras que los otros 2 nodos destacados, `192.168.1.11` y `192.168.1.101` ni siquiera existen en la red, las direcciones no fueron asignadas y ningún nodo las está usando. Sólo podemos suponer que este host tiene un software mal configurado o defectuoso. Las interacciones de los nodos se pueden ver en más detalle, en el siguiente gráfico donde se puede ver a los nodos destacados y las relaciones con los otros nodos. Notar que en el gráfico, así como en el experimento, también aparece un nodo con dirección `169.254.255.255` la cual es la dirección que el sistema operativo windows asignó a un nodo cuando este no puede contactar a ningún servidor DHCP para que le asigne una dirección libre. Usualmente los nodos operan con esta dirección durante segundos o minutos, hasta que se pueda proveer de una dirección válida.

Figura 3: Visualización red hogareña, en amarillo los nodos destacados



El gráfico confirma que los nodos que se destacaron, son solo accedidos por un nodo.

4. Red laboral

Para explorar nuevas topologias y relaciones se decidió monitorear una red laboral en una oficina de mediana envergadura. En esta red local conviven computadoras de empleados, servidores de distinto uso, impresoras de red y otros dispositivos de oficina.

El primer experimento, en el cual se ve la red como una fuente con 2 símbolos unicamente (Unicast y Broadcast), monitoreó la red local a través de un enlace Ethernet durante 20 minutos. Como era de esperarse cerca del 100 % del tráfico es unicast dado que es una red con mucho tráfico hacia un servidor en particular o hacia el gateway.

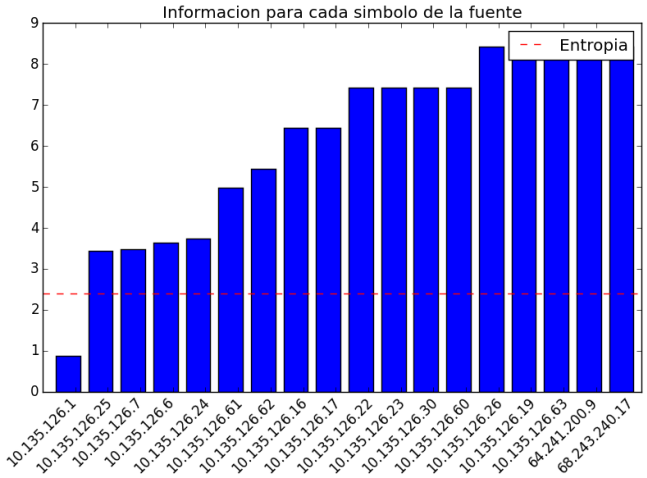
Figura 4: Información de S, Red laboral

	frecuencia	información
$S_{broadcast}$	0.03	5.06
$S_{unicast}$	0.97	0.04

Estos valores nos dan una entropía de 0.19 bits (siendo el máximo 1 dado que hay 2 simbolos en principio equiprobables), dado que el tráfico broadcast sin ser despreciable sólo es un porcentaje pequeño del tráfico total.

En el segundo experimento, el cual modela la red basado en la dirección a resolver en mensajes ARP, usamos la misma captura utilizada en el experimento anterior. Bajo nuestra definición para destacar nodos (que la información del simbolo de ese nodo sea menor o igual a la entropía de la fuente) sólo un nodo fue destacado por amplio margen, el gateway por defecto de la red.

Figura 5: Información red laboral

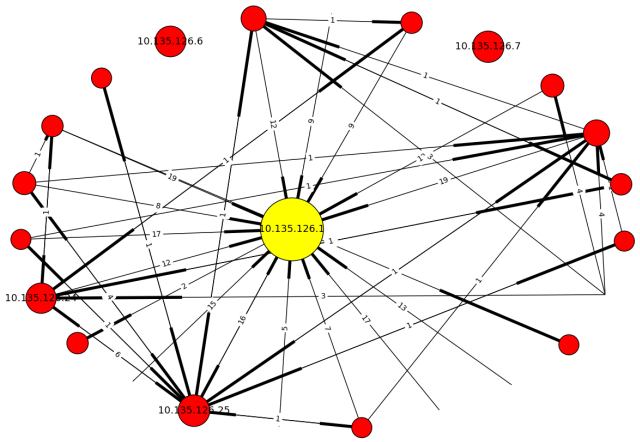


La visualización de la red modelada como la fuente descriptiva muestra una topologia y relación entre nodos muy similar a la realidad en términos de trafico total y disposición.

Un análisis manual del tráfico evidenció dos fenómenos ARP no usuales, ARP gratuitos y ARP de sondeo. Los ARP gratituitos pueden ser tanto `who-has` como `is-at` donde tanto la dirección de origen como la dirección de destino son las mismas (y la dirección de enlace destino es broadcast). El uso de los mismos es precargar o refrescar las tablas de otros nodos para evitar tener que traducir en tiempo real una dirección de red. Los ARP de sondeo tienen un fin similar, un nodo al que se le asignó manualmente o automaticamente una dirección de red, envia un ARP con dicha dirección a la red y si algún nodo le responde entonces sabrá que la dirección ya está en uso y evitará así la colisión de alguna forma. Ninguno de estos tipos de mensajes están oficialmente documentados pero forman parte de toda red que necesite

autoregularse correctamente.

Figura 6: Visualización red laboral, en amarillo los nodos destacados



5. Red pública

Situados en un patio de comidas, monitoreamos la red wifi de un shopping importante con el objetivo de analizar una red con muy poco control y en constante cambio.

El primer experimento, el cual ve la red como una fuente de 2 símbolos unicamente, resulto en números similares a la red hogareña aunque en un principio se podría suponer que el tráfico broadcast iba a ser mucho más elevado con tantos nodos entrando y saliendo de la red.

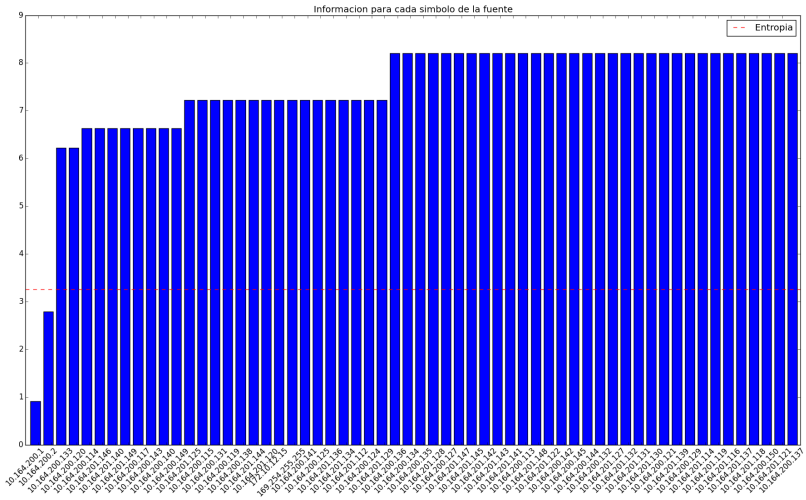
Figura 7: Información de S pública

	frecuencia	información
$S_{broadcast}$	0.21	2.25
$S_{unicast}$	0.79	0.34

Estos valores dan una entropía de 0.65 bits (siendo el máximo 1 dado que hay 2 símbolos, en principio equiprobables). Suponemos que la similitud con la red hogareña está en que el tiempo en que monitoreamos la red es menor o igual que el tiempo de cada nodo en la red.

En el segundo experimento, el cual modela la red basado en la dirección a resolver con mensajes ARP, usamos la misma captura utilizada en el experimento anterior. Nuestra herramienta y definición de nodo destacado encontró 2 nodos a destacar: uno se sabe que es el gateway por default, el segundo nodo no se sabe con precisión , posiblemente sea otro gateway para otro sector o para diferentes clientes (dado que analizando la dirección de enlace podemos ver que es un router de reconocida marca)

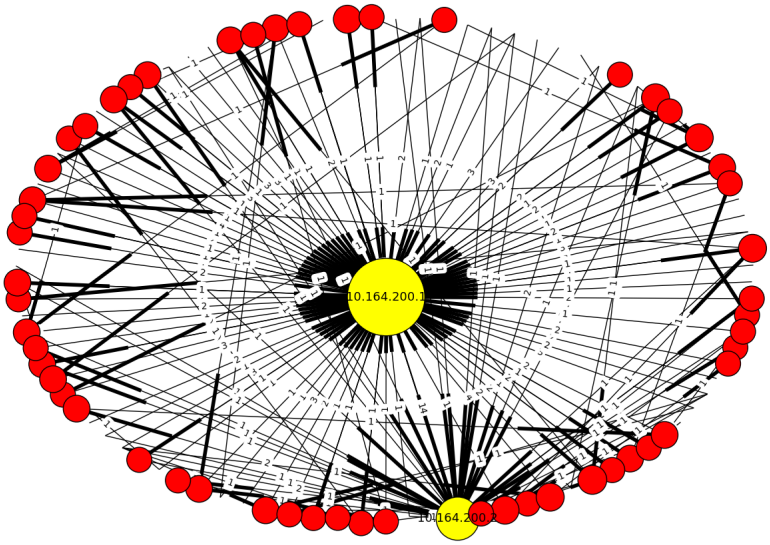
Figura 8: Información red pública



Se pueden notar 3 grupos de nodos basados en la información: destacados, aquellos que fueron accedidos dentro de la red por otros nodos y una gran cola de nodos que solo figuran por un otro ARP capturado. Lo mismo se refleja en la visualización de la red a partir de la fuente S1 donde se puede ver como los nodos destacados son varias magnitudes más grandes que el resto.

Al igual que la red laboral antes descripta, volvimos a observar ARP gratuitos y ARP de sondeo provenientes del gateway y de algunos nodos.

Figura 9: Visualización red pública, en amarillo los nodos destacados



6. Red hogareña casa Ale

En ésta sección se analizará una red hogareña pero del hogar de otro integrante. El objetivo en este caso, es mostrar una red mucho más chica, con un grafo con menor cantidad de nodos (hosts) y ver si la cantidad de paquetes Unicast y Broadcast son similares o no. Para este caso se espera que los valores en promedio de paquetes Unicast y Broadcast sean parecidos al experimento realizado en el hogar del otro integrante (punto 1) red hogareña), ya que en un hogar, cada dispositivo móvil y cada computadora realizan sus propias tareas, los usuarios de cada dispositivo realizan sus propias actividades y por lo general, no están relacionadas entre sí, por lo tanto, la cantidad de paquetes Unicast debería ser mayor, claramente, a la cantidad de paquetes Broadcast.

Dada la fuente binaria S, donde S se define como S = ($S_{unicast}$, $S_{broadcast}$), tomando como la entropía máxima = 1, suponiendo que los eventos son equiprobables, a continuación se mostrarán los calculos de la entropía y los porcentajes de paquetes Unicast vs Broadcast:

Figura 10: Información de S pública

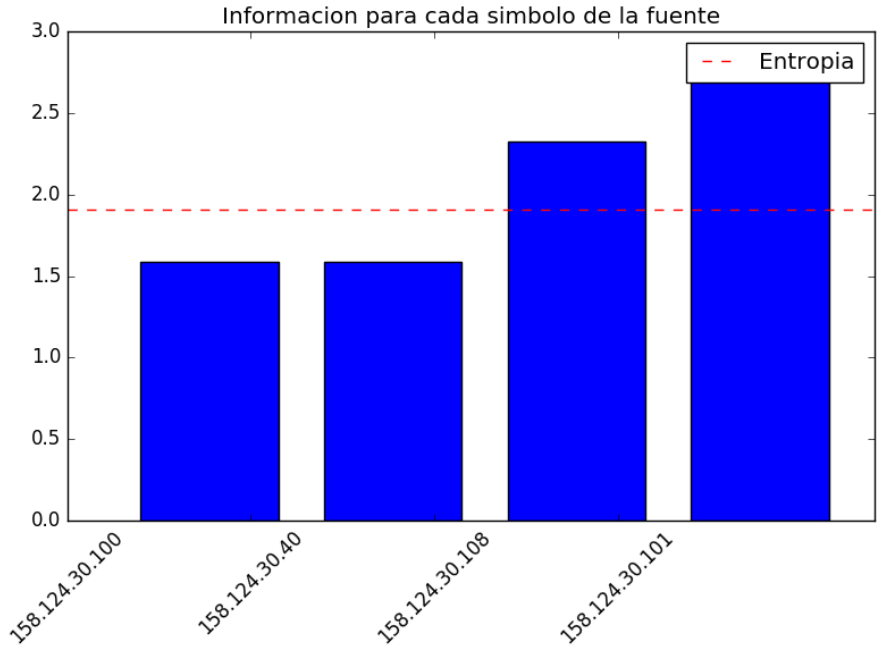
	frecuencia	información
$S_{broadcast}$	0	0
$S_{unicast}$	0	0

Como se puede observar en la siguiente tabla, hay dos nodos que tienen valores que están por debajo de la entropía media, el host 158.124.30.40 y el host 158.124.30.100. El primero se debe a la computadora de escritorio que está conectada al router y el segundo se debe a uno de los celulares que más actividad

en internet estaba teniendo al momento de la captura de paquetes. El resto de los nodos de deben al resto de los dispositivos móviles.

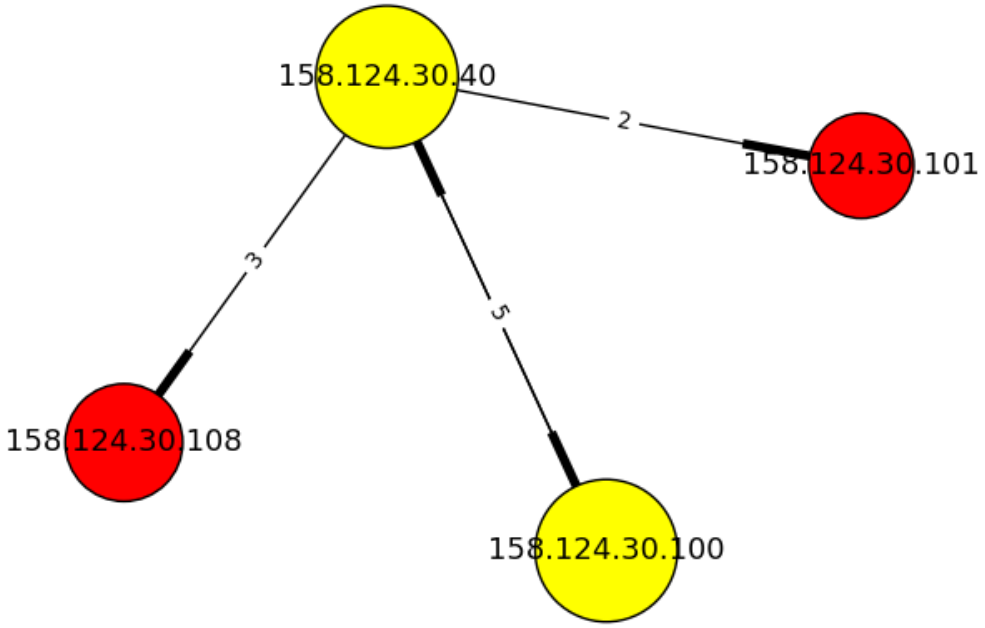
A continuación se podrá obsevar la actividad de cada nodo, en cuanto a cantidad de información con respecto a la fuente S1, que según como se definió anteriormente, es la cantidad de nodos distintos que aparecen en la red. Estos datos se comparan contra la entropía media que es la línea roja que aparece en el gráfico.

Figura 11: Información red pública



En el siguiente gráfico se mostrará el grafo subyacente de mensajes ARP que se formó con los paquetes ARP que estaban circulando en la red que observamos. Como se descubrió antes, aparecen en color amarillo, los nodos destacados, es decir, que sus valores de información están por debajo de la entropía media. Se pueden observar cuatro nodos, un nodo destacado que es el principal (el 158.124.30.40), esa ip se corresponde con la computadora de escritorio que esta conectada por cable de red al router, este puede ser un motivo por el cual, los nodos restantes estén enlazados a este, en el grafo. Los nodos restantes se corresponden a telefonos móviles, esto se pudo averiguar chequeando la ip de cada celular. Por lo tanto, en este caso, no se detectaron nodos anómalos.

Figura 12: Visualización red pública, en amarillo los nodos destacados



7. Conclusiones

Luego de analizar cuatro redes con configuraciones diferentes podemos concluir que el tráfico unicast parece ser usualmente superior al broadcast si se monitorea una red durante el tiempo suficiente, tal vez por que las cuatro redes tienen un proposito común que es comunicar una red local al exterior u a otra red. Además la frecuencia de los paquetes broadcast aumenta considerablemente para el caso de la red pública, ya que es más probable que se quiera transmitir un mismo paquete para todos los hosts.

También podemos mencionar que nuestra elección de nodos distinguidos resultó ser muy útil, no solo para distinguir el usual gateway por default sino como vimos también para detectar otros gateways no asignados al nodo monitor o bien para diagnosticar nodos que no funcionan bien, como vimos en el primer caso. Esta elección se valió exclusivamente en la entropía de la fuente, la cual sirvió efectivamente como un límite al cual un símbolo/nodo debe acercarse para ser relevante en la red y que la entropiá parece ser mayor cuando mas cantidad de nodos hay en la red, esto tiene sentido si se modela usando ARP dado que vamos a incrementar la cantidad de símbolos con poca frencuencia en nuestra fuente.

Cabe destacar también el descubrimiento de técnicas utilizadas por routers y nodos para comunicar datos a través de ARP con el objetivo de mejorar la performance de la red, utilizando mensajes no requeridos por ningún otro nodo.